

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-187497

(43)Date of publication of application : 04.07.2000

(51)Int.Cl.

G10L 15/22

G10L 15/28

(21)Application number : 10-364066

(71)Applicant : HITACHI LTD
HITACHI ULSI SYSTEMS CO LTD

(22)Date of filing : 22.12.1998

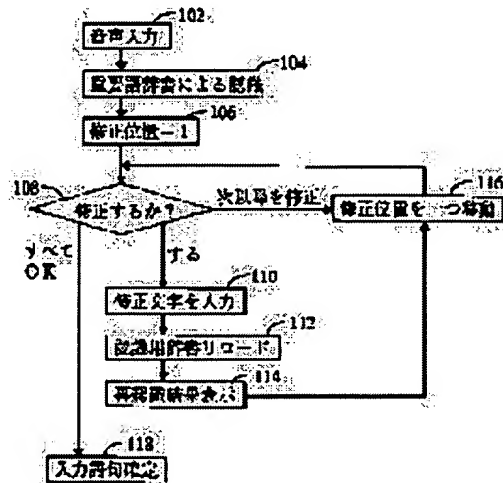
(72)Inventor : OBUCHI YASUNARI
KOIZUMI ATSUKO
TANAKA MAKOTO

(54) LANGUAGE INPUT DEVICE BY VOICE

(57)Abstract:

PROBLEM TO BE SOLVED: To surely recognize a frequently used word and to recognize even an infrequently used word without consuming a useless calculation time by recognizing a significant word in a first step and simultaneously performing a correction work of a recognition error and the load of an insignificant word dictionary.

SOLUTION: A user inputs words and phrases to be inputted by voices (S 102). The inputted voices are recognized by using the significant words containing only the significant words and phrases among the words and phrases incorporated in a large vocabulary dictionary (S 104). The user specifies the presence of the correction in the state that the recognition result and a corrective position are displayed (S 108). When though the characters displayed as the corrective position are suited, the errors exist in the characters after them, the corrective position is moved (S 116). When the characters displayed as the corrective position are mistaken, the correct characters are inputted (S 110). Only the words and phrases coinciding with the decided characters among the recognition objective words and phrases are extracted to be loaded as a new recognizing dictionary (S 112).



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-187497

(P2000-187497A)

(43) 公開日 平成12年7月4日(2000.7.4)

(51) IntCl.

識別記号

F I

テーマコード(参考)

G 1 0 L 15/22
15/28

G 1 0 L 3/00

5 7 1 V 5 D 0 1 5
5 7 1 H
5 6 1 C
5 6 1 E

審査請求 未請求 請求項の数11 O L (全 5 頁)

(21) 出願番号 特願平10-364066

(22) 出願日 平成10年12月22日(1998.12.22)

(71) 出願人 000005108

株式会社日立製作所
東京都千代田区神田駿河台四丁目6番地

(71) 出願人 000233169

株式会社日立超エル・エス・アイ・システムズ
東京都小平市上水本町5丁目22番1号

(72) 発明者 大淵 康成

東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(74) 代理人 100068504

弁理士 小川 勝男

最終頁に続く

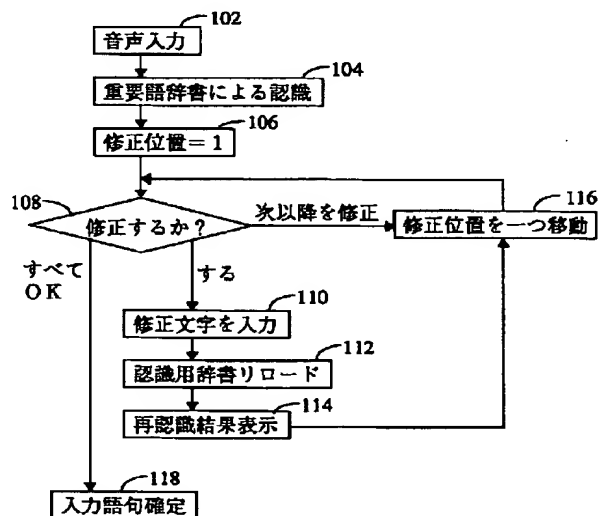
(54) 【発明の名称】 音声による言語入力装置

(57) 【要約】 (修正有)

【課題】 多数の語句を含む辞書を用いる音声認識による入力において、計算時間を短縮し認識精度を向上させる。

【解決手段】 最初の認識は、重要語辞書のみを用いて行なう。入力音声に対応する単語が重要語辞書に含まれていない場合には、最初の認識結果が正解にならないという問題が生じるが、認識結果に修正を加える段階で非重要語も用いて再度認識を行なう。このようなシステムにより、ユーザの手間を最小限に押さえた入力装置が実現可能となる。

図 1



【特許請求の範囲】

【請求項 1】あらかじめ用意された認識対象語句リストをもとに、入力音声を認識する装置において、認識対象語句のうち、その一部のみを認識処理の対象として限定する手段と、

入力音声、もしくは入力音声から抽出した特徴量を記憶しておく手段と、

前段階での認識結果をユーザが部分的に修正することを可能にする手段と、

前段階での認識結果の修正をもとに、前段階では使用されなかった認識対象語句をあらたに認識処理の対象として追加する手段と、

追加された認識対象語句に対し、記憶されている入力音声、もしくは入力音声から抽出した特徴量をもとに、認識処理を実行する手段と、

を有することを特徴とする音声による言語入力装置。

【請求項 2】認識対象語句を限定する手段において、ユーザが過去に入力したことのある語句を一定数だけ含むようにするための手段を有することを特徴とする、請求項 1 記載の音声による言語入力装置。

【請求項 3】認識対象語句を限定する手段において、各認識対象語句に重要度に相当する値を付与する手段と、

ユーザが過去に入力したことのある語句についてはこの重要度に一定の値を加算もしくは乗算した値を付与する手段と、

両者の値を比較して、限定すべき認識対象語句を決定する手段と、

を有することを特徴とする、請求項 2 記載の音声による言語入力装置。

【請求項 4】認識対象語句を限定する手段において、限定する語句のリストを複数保持しておき、状況に応じて使い分けるための手段を有することを特徴とする、請求項 1 記載の音声による言語入力装置。

【請求項 5】認識結果をユーザが修正することを可能にする手段において、

認識結果、もしくは認識結果に何らかの修正を施した結果を文字で表示する手段と、

表示された文字の中で、修正したい文字を特定するための手段と、

修正したい文字が変わって、正しい文字を入力する手段と、

を有することを特徴とする、請求項 1 記載の音声による言語入力装置。

【請求項 6】修正したい文字を特定するための手段において、

認識結果の各文字の信頼度を推定する手段と、

各文字の信頼度の推定値に基づき、ユーザの負担を最も軽くするために最適な修正位置を推定する手段と、

推定された修正位置をユーザに示す手段と、

を有することを特徴とする、請求項 5 記載の音声による言語入力装置。

【請求項 7】正しい文字を入力する手段において、文字を音声で読み上げることによって入力することを可能にする手段を有することを特徴とする、請求項 5 記載の音声による言語入力装置。

【請求項 8】正しい文字を入力する手段において、文字に関連する単語を音声で読み上げることによって入力することを可能にする手段を有することを特徴とする、請求項 5 記載の音声による言語入力装置。

【請求項 9】正しい文字を入力する手段において、現在表示されている文字に代わって入力される可能性の高い文字を推定する手段と、

推定された候補文字を表示する手段と、

表示された候補の中からユーザが正しい文字を選択することを可能にする手段と、

を有することを特徴とする、請求項 5 記載の音声による言語入力装置。

【請求項 10】候補の中から正しい文字を選択する手段において、正しい文字に該当する番号を音声で読み上げることによって入力することを可能にする手段を有することを特徴とする、請求項 9 記載の音声による言語入力装置。

【請求項 11】認識結果の修正をもとに、認識対象語句を追加する手段において、認識結果に対して、最後の修正が行なわれた位置、もしくは次の修正が行なわれようとしている位置を検出する手段と、

検出された位置よりも前の部分の内容を特定し、その内容に矛盾しない認識対象語句のみを追加することを可能にする手段と、

を有することを特徴とする、請求項 1 記載の音声による言語入力装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、多数の語句から成る辞書が設定されている状態で、音声を入力することによってその辞書に含まれる任意の語句を特定し、さらに入力を試みた語句が辞書に含まれない場合には含まれないという結果を出力するために用いる、音声認識装置に関する。

【0002】

【従来の技術】音声認識装置は、入力された音声を非常に短い時間ごとに切りわけ、信号処理の手法に基づき分析し、いくつかの特徴量の組に変換する。一方、認識対象となるすべての語句は、それぞれ対応する特徴量の組から成る標準パターンに変換されて保持されている。それぞれの標準パターンは入力音声から得られた特徴量と比較され、類似度をスコアとして保持する。入力が進むにつれてこのスコアが累積され、入力が終了した時点

で最大スコアを持つ標準パターンに対応する語句が認識結果として出力される。

【0003】

【発明が解決しようとする課題】従来の音声認識装置では、すべての認識対象語句の標準パターンに対する類似度を計算しなければならないため、認識対象語句の数が増えるにつれ、処理量もそれに比例して増大するという問題が生じる。例えば、市販されている国語辞典の収録単語すべてを認識しようとする、5万から10万単語程度の認識を行わなければならないが、これを安価な計算機でリアルタイムで実行することは非常に困難である。また、認識対象語句の数が増えるにつれ、似通った音を持つ語句の組み合わせも増えることから、認識誤りも生じやすくなる。特に、めったに使われない語句の存在によって、非常によく使う語句の認識率までもが低くなってしまったような弊害が生じることもある。本発明の目的は、このような状況において、よく使われる単語を確実に認識するとともに、あまり使われない単語であっても無駄な計算時間を消費することなく認識することができるような装置を提供することである。

【0004】

【課題を解決するための手段】本発明においては、最初の認識実行時には重要語のみを認識対象とすることによって、計算時間と認識精度の両者を適正な範囲に収めることができる。ただし、ユーザにとっては自分の入力しようとしている単語が重要語に含まれているかどうかを知ることができないので、非重要語を入力しようとした場合の救済手段を用意しておく必要がある。一方、音声認識では多少なりとも認識誤りが生じることが避けられないため、これを修正する手段が必要となる。そこで、非重要語が入力された場合には、認識誤りを修正する作業のあいだに認識対象語句を入れ替え、非重要語も含むようにする。これによって、ユーザから見ると、非重要語であったが故に一回目の認識で認識誤りが生じただけのように見え、わずかな修正作業だけであたかも最初からすべての語句を候補とした認識を行なったかのような結果が得られることになる。

【0005】

【発明の実施の形態】以下、図を用いて本発明の実施例を説明する。本発明の音声による言語入力装置は、一般的な装置構成、すなわち、マイク等の音声入力装置を含み、マウス、タッチパネルまたは入力ボタンなどの入力装置、認識の途中結果や最終結果を出力するスピーカや表示装置などの出力装置、入力された音声処理（後述のプログラムを実行）する処理装置、処理装置に読み出されて認識処理を行うプログラム、処理プログラムの実行に伴い必要な音声特徴リスト、認識辞書、認識途中の及び最終結果等を記憶するメモリ等を有している。本発明では、処理の手順に特徴を有している。

【0006】図1は、本発明を用いた音声認識装置の一

実施例における処理の概要を表わしている。ユーザはまず初めに、自分の入力したい語句を音声によって入力する（102）。入力された音声は、大語彙辞書に含まれる語句のうち重要な語句だけを含む重要語辞書を用いて認識される（104）。このとき、入力された音声波形、もしくはそこから抽出された特徴量は記憶装置に保持される。認識結果は画面に表示され、カーソルもしくは反転表示などの手段で修正位置が示される。最初の認識の直後には修正位置は1文字目とすることが標準的であるが、認識結果の1文字目の信頼度が非常に高いときには修正位置を2文字目として開始するなど、各文字の信頼度に基づき修正モードを開始することも可能である。認識結果と修正位置が表示された状態で、ユーザは修正の有無を指定する（108）。認識結果が入力しようとした語句と全く違う場合には、すべてOKであるということを、ボタンもしくは音声などによって入力することにより、入力語句を確定させる（118）。修正位置として表示されている文字はあっているが、それより後の文字に誤りがある場合には、修正位置を移動させるボタンを押す、もしくは修正位置の移動に対応した音声を入力することなどにより、修正位置を移動させる（116）。修正位置を移動させると、再びユーザの指示を待つ（108）状態に戻る。修正位置として表示されている文字が誤っている場合には、正しい文字を音声等で入力する（110）。入力された文字が確定した段階で、その文字より手前の文字はすべて確定したと見なすことができる。そこで、すべての認識対象語句のうち、既に確定している文字と合致するもののみを抽出し、新たな認識用辞書としてロードする（112）。先頭からの何文字かが限定されるため、新たにロードされる辞書は認識対象語句全体に比べて非常に小さいものになるため、計算時間と認識精度の両面で高い性能を出しうるものとなる。こうしてロードした辞書をもとに、最初の認識（104）のときに保持されていた入力音声の特徴量を使って再び音声認識を行ない、認識結果を表示する（114）。このとき、修正モードで確定された部分が正解と共通であることは明らかなので、修正位置を一つだけずらし（116）、再び認識結果の修正モードに入る（108）。以下、同様の作業を繰り返し、入力語句と表示が一致した段階で入力語句確定（118）となる。

【0007】図2は、重要語辞書の作成方法の一例を表わしている。重要語辞書の基本重要語領域（202）は、あらかじめ指定された一般的な重要語を列挙したもので、常に用いられる。一方、ユーザが一度入力した単語は再び用いられることが多いので、これらの単語を重要語辞書に含んでおくことが望ましい場合がある。このような場合、ユーザ辞書領域（204）を確保しておき、ユーザが入力した単語を順次登録する。この例では、ユーザが「にんしき」と入力したのを受けて、その

単語がユーザ辞書領域の先頭に登録される。このとき、ユーザ辞書領域の先頭にあった「おんせい」という単語は一つ下に移り、以下すべての単語がひとつずつシフトする。最も下に登録されていた単語（この例でいうと「じしょ」という単語）は、ユーザ辞書領域から抹消される。ここで述べた方法は最も単純なものであるが、この他に、各単語に重要度のスコアを付けておき、使用頻度とスコアの両者を懸案してユーザ辞書領域の登録単語を決めるといった方法も可能である。

【0008】図3は、最初の音声認識を終了し、修正モードに移った直後の画面の例である。表示画面（302）には、ユーザの発声に対する認識結果が表示され、修正位置がカーソル（304）によって示されている。例えばユーザが「せんぶうき」と発声し、図のように表示された場合には、認識結果を修正する必要がないため、認識確定ボタン（306）を押すことによって入力を確定させる。認識確定ボタンを用いる代わりに、「にんしきかくてい」などの音声によって入力を確定させることもできる。次に、ユーザが、例えば「せいひょうき」と発声して、図のように表示された場合を考える。この場合、一文字目の「せ」は正しいので、カーソル移動ボタン（308）を押すことによって、修正位置を右に一文字分移動させる。カーソル移動ボタンを用いる代わりに、「カーソルいどう」などの音声によって修正位置を移動させることもできる。修正位置が二文字目となった段階で、本来入力したい単語の二文字目は「い」であるので、ユーザは「い」と発声する。この発声が認識されると、二文字目の表示が「い」に変化する。ただし、単音節の音声認識性能が低い場合においては、この「い」という音声の認識に対し、結果を一文字もしくは複数候補表示し、ボタン操作等による選択、確定の動作がなされた段階で二文字目の「い」を確定させることにしてもよい。こうして二文字目が確定した段階で、すべての認識対象単語の中から、「せい」で始まる単語のみが抽出され、それらを対象にして、最初に入力した音声

分を含む単語を抽出しようとして、そのような単語が一つも登録されていない場合には、その旨を画面に表示することによって、ユーザはその単語が認識対象として登録されていないことを知ることができる。

【0009】図4は、修正モードの画面の別の実施形態を表わす図である。仮に修正位置が二文字目であるとして、カーソルをその位置に移動させた状態で、候補文字表示ボタン（402）を押すことによって、候補文字表示ウィンドウ（404）を表示させる。候補文字表示ボタンを用いる代わりに、「こうほもじひょうじ」などの音声によって候補文字表示ウィンドウを表示させることもできる。候補文字表示ウィンドウには、修正のために置き換えられる可能性の高い候補文字がいくつか表示される。必要であれば、それらの文字には番号を付与させることもできる。ユーザは、候補文字選択ボタン（406）を使ってこれらの中からある文字を選ぶか、変換したい文字を音声で入力するか、もしくは変換したい文字に付与されている番号を音声で入力するなどの手段によって変換する文字を確定する。

【0010】

【発明の効果】数万から数十万にのぼる語句を含む辞書に対する音声入力は、計算時間と認識精度の両面で非常に困難である。しかしながら、そのような大語彙辞書に対する入力を、キーボードなどの操作を用いずに簡便に行ないたいという需要は大きい。本発明では、重要語に対する認識を第一段階で行なうことによって、重要語に対する入力の手間を最小限に抑え、同時に非重要語の存在に起因する認識率劣化を防止する。その一方で、認識誤りの修正作業と非重要語辞書のロードを同時に行なうことにより、非重要語の入力に際して重要語に比べて著しい不便が生じることを避け、ユーザはあたかも最初からすべての語句に対する入力が自然に行なわれているように感じることができる。

【図面の簡単な説明】

【図1】本発明による装置における処理の概要。

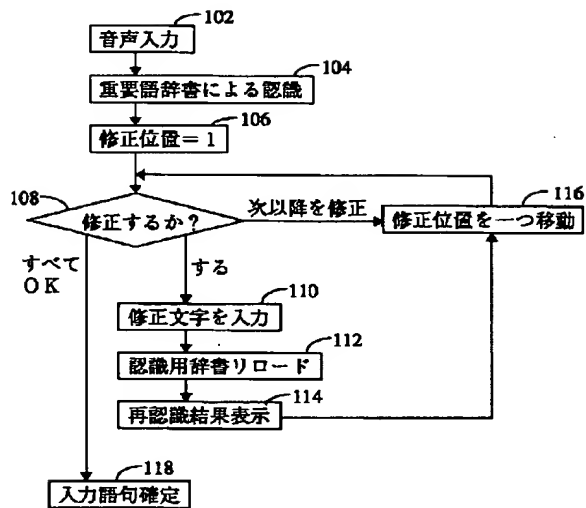
【図2】本発明で用いる重要語辞書の作成例。

【図3】本発明による装置の修正モードにおける画面出力の例（1）。

【図4】本発明による装置の修正モードにおける画面出力の例（2）。

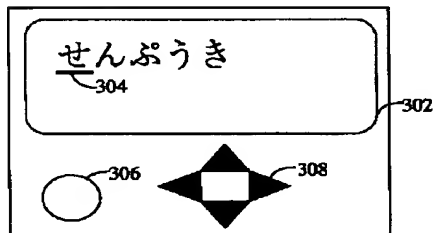
【図1】

図1



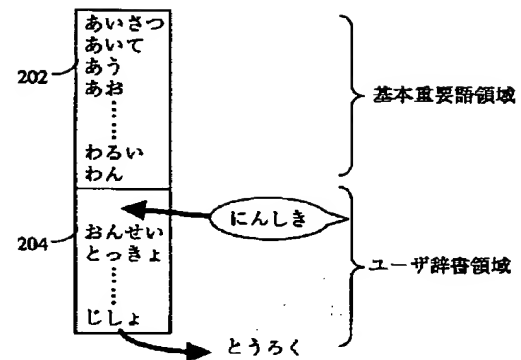
【図3】

図3



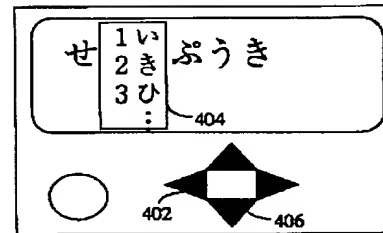
【図2】

図2



【図4】

図4



フロントページの続き

(72)発明者 小泉 敦子
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 田中 誠
東京都小平市上水本町5丁目20番1号 株
式会社日立超エル・エス・アイ・システム
ズ内

Fターム(参考) 5D015 LL04 LL05 LL09 LL10

THIS PAGE BLANK (USPTO)